Circadian control of the sleep–wake cycle

Domien G.M. Beersma⁎, Marijke C.M. Gordijn

Department of Chronobiology, University of Groningen, PO Box 14, 9750 AA, Haren, The Netherlands

Received 14 August 2006; accepted 4 September 2006

Abstract

It is beyond doubt that the timing of sleep is under control of the circadian pacemaker. Humans are a diurnal species; they sleep mostly at night, and they do so at approximately 24-h intervals. If they do not adhere to this general pattern, for instance when working night shifts or when travelling across time zones, they experience the stubborn influence of their circadian clock.

In recent years much has been discovered about the organisation of the circadian clock. New photoreceptor cells in the retina have been found to influence the input to the clock, and much of the molecular machinery of the clock has been unravelled. It is now known that the circadian rhythm of sleep and wakefulness is only loosely coupled to the circadian rhythm of the pacemaker. New theories have been proposed for the functions of sleep and the sites at which those functions are executed. In spite of this rapid increase in knowledge of the circadian clock and of sleep regulatory processes, much remains to be discovered concerning the precise interaction between the biological clock and sleep timing. This is particularly unfortunate in view of the 24-h demands of our society for 7 days a week. Too little is known about the negative consequences of the societal pressures on well-being and performance.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Circadian rhythm; Sleep regulation; Mathematical models; Two-process model; Opponent processes model

1. Introduction

In most mammals other than humans, the circadian control of the sleep–wake cycle is so overwhelmingly present that the sleep–wake pattern is generally used to monitor the behavior of the circadian pacemaker. Yet, there is more to the regulation of activity and rest than just the influence of the circadian pacemaker. All kinds of external and internal influences, like environmental light intensity, environmental temperature, stress, presence of other animals, and health condition have their own influence on the final pattern of activity and rest. In humans it is well accepted that the circadian pacemaker in the suprachiasmatic nucleus (SCN) of the hypothalamus exerts gentle control over the sleep–wake cycle, as is demonstrated by the extreme situations by which we can nap during the day or stay awake at night, for instance. Commonly, the circadian pacemaker is not capable of adjusting to such rapid alterations of the sleep–wake pattern [1]. Since the alternation between sleep and wakefulness has its own impact on a wide variety of bodily functions, this implies that two kinds of 24-h rhythms exist: circadian rhythms controlled by the endogenous circadian pacemaker, and 24-h rhythms resulting from the alternation between sleep and wakefulness.

The distinction between sleep-related 24-h rhythms and 24-h rhythms related to the circadian pacemaker is exploited in so-called forced-desynchrony experiments and to a lesser extent in constant-routine experiments. In forced-desynchrony experiments, sleep and wakefulness are forced to alternate at a frequency different from that of the circadian pacemaker [2]. By doing so, it becomes possible to distinguish circadian variation in a variable caused by sleep-related processes from the variation related to the circadian pacemaker [3,4]. Such experiments using forced-desynchrony have, for instance, demonstrated that the duration of REM sleep episodes depends on circadian phase (maximum duration shortly after the minimum of core body temperature) and on prior sleep duration (longer REM sleep episode duration after longer prior sleep) [3]. In constant-routine experiments, subjects stay awake for more than 24 h, in constant environmental temperature and light exposure conditions, while staying in a semi-recumbent position and taking similar snacks at frequent intervals. In such experiments the amount of
circadian variation detected in any variable of interest is attributed to the influence of the circadian pacemaker, since other circadian influences are kept constant [5].

Models have been developed to try and explain how the circadian pacemaker influences 24-h patterns of behavior. Examples are the two-process model of sleep regulation [6–8] and the opponent processes model [9]. Such models distinguish between processes that are almost entirely dependent of behavioral state, like the need for sleep, and processes that are almost entirely independent of behavioral state, like the circadian clock signal provided by the circadian pacemaker [10].

The models of sleep regulation are based on certain concepts regarding the interaction between the processes involved. In the opponent processes model [9], it was originally proposed on the basis of data of SCN lesions in squirrel monkeys that the increasing need for sleep during waking is counteracted by a circadian process that increasingly stimulates wakefulness during daytime for diurnal species (see Fig. 1). For a comprehensive critical review of the opponent process model see [11]. Later studies in humans [3,12,13] generalised the concept by proposing that the progressive decrease in the need for sleep during sleep is similarly mirrored by a circadian process that increasingly stimulates sleep in the other half of the nycthemeron. The opponent processes model is a conceptual model that has not been elaborated into a mathematical formulation. Consequently, the quantitative predictions needed for experimental testing are not available. For the two-process model of sleep regulation a different situation applies. Here a mathematical translation of the concept was put forward right from the beginning [7], and this has been improved regularly [8,14–16]. Conceptually, the model considers the alternation of wakefulness and sleep to result from the interaction of two processes, S and C (see Fig. 2). Process S represents sleep need. It increases during waking and decreases during sleep. Functionally this implies that sleep would serve a recovery function. Process C is entirely determined by the temporal sequence of behavioral states. Process C, in contrast, is totally controlled by the circadian pacemaker, irrespective of behavioral state, and is proposed to set limits to process S. Those limits vary with time of day. As soon as S reaches the lower limit during sleep, subjects will wake up. If S reaches the upper limit during waking, sleep will be initiated. It has been demonstrated that the activity in the low frequency range (1–4 Hz) in the electroencephalogram of non-REM sleep behaves as predicted for a measure representing the decline of process S during sleep [6,17,18] and its increase during waking [19,20]. This relationship alone turned out to be sufficient to quantify the dynamical properties of both processes [7]. As a result, the consequences of various sleep–wake protocols for sleep need and for timing of subsequent sleep can be calculated, for instance in shift work.

The two-process model of sleep regulation has been modified for other purposes. A major effort has been paid to predict subjective ratings of sleepiness and performance. For that purpose, the dynamic properties of the relevant processes had to be modified and a third process was added accounting for ‘sleep inertia’, i.e. the time it takes after awakening to become fully alert [21–23].

2. Model limitations

Models are always simplifications of reality. Mostly this is unavoidable because the real-life situation is so complex that it is impossible to take account of every aspect of regulation. The simplifications are at the same time advantageous and disadvantageous. They are advantageous because they help to understand major principles of regulation, and they are disadvantageous because other major principles may be overlooked [24]. A shortcoming of the two-process model of sleep regulation, for instance, is that it is deterministic. Although it is acknowledged that the thresholds to process S may fluctuate stochastically, it is the event of S reaching the upper threshold that triggers sleep initiation. After the switch, according to the
model, S has to travel all the way to the wake threshold before sleep is terminated again. The occurrence of short awakenings at night cannot be understood with the model, nor can the occurrence of short naps during the day. Qualitatively, the opponent processes model [12] does provide explanations for the short waking bouts during sleep and the short duration of naps: according to the model, the circadian pacemaker draws the system towards wakefulness during the active period and pushes it towards sleep in the inactive interval. Intervals of waking during the sleep period are likely to end soon due to the influence of the pacemaker. The same applies to the naps: after initiation they are apt to last only for a short time because of the opposing influence of the pacemaker. Yet, neither one of the two models explains why intermittent waking bouts occur during sleep and why naps occur during the day.

The question as to how the pacemaker interacts with homeostatic processes such as process S is important to answer. The answers would give insight in the problems of jet lag, of shift work, and for instance, in the problems that early and late chronotypes have with the timing of their sleep.

Another shortcoming of the models is that the two processes involved are largely theoretical constructs. The need for sleep, expressed in process S, does not refer to a specific quantifiable variable of a specific physiological process. Even though SWA in non-REM sleep and slow eye movement density during waking would provide us with reliable estimates of the current state of process S, these variables themselves result from complex processes, the essential aspects of which that contribute to the need for sleep are not known. The same applies to process C. There is no doubt that a circadian pacemaker influences all kinds of processes in the body, but where and how this pacemaker influences the timing of sleep and wakefulness is not yet clear.

3. New developments

3.1. Process S

Process S is thought to represent the “need for sleep”, but what, exactly, is this need for sleep? Also, extension of an interval of wakefulness makes it increasingly harder to resist sleep, but why is this? What are the essential physiological processes in the brain or in the body that require sleep? It has been proposed that aspects of the immune system would be at the basis of the need for sleep, because sleep deprivation affects the immune response to a wide variety of infections as well as the response to vaccinations [25]. However, a thorough recent review of the literature does not convincingly demonstrate a major role for sleep in immune functions [26].

Similar conclusions apply to some important aspects of alertness regulation, including glycogen metabolism, adenosine metabolism, and hypocretin metabolism. Glycogen is considered to form an energy reserve for the brain, to be utilized at times of shortage of glucose. Benington and Heller [27] suggested that the brain would run out of glycogen during wakefulness and that non-REM sleep in particular was required for its replenishment. However, the rapid increase of glycogen during a short interval of slow wave sleep [28], the continuation of glycogen resynthesis during the waking state and the distribution of glycogen in the brain [29,30] argue against a major role for sleep in glycogen regulation. Adenosine regulation seems related to sleep regulation much more closely [31]. However, the regulation of adenosine is not monotonously dependent on sleep duration or on wake duration (as is expected for a variable underlying process S), and it also may depend on circadian phase [32]. In terms of the two-process model, this means that not all of the 24-h variation in adenosine concentration is related to sleep, but that part of it is controlled by the circadian pacemaker. Hence, the finding suggests that adenosine becomes involved in circadian regulation after the interaction between process S and process C has taken place. Similar conclusions can be drawn for the regulation of hypocretin, since Zeitzer et al. [33] demonstrated that hypocretin shows both circadian and homeostatic variation.

At the psychological level, sleep is thought to be needed for such functions as memory consolidation [34]. Although this is a very interesting proposition, it is not easy to prove unequivocally a direct connection between memory consolidation and sleep. It may well be that processes profiting from sleep (such as alertness) positively influence recall, and that sleep itself is only indirectly involved. For proper testing of the hypothesis it would be desirable to know the relevant substrate for memory consolidation. Recent studies, reviewed by Tononi and Cirelli [35], might provide the relevant information. Those authors have shown that synaptic connectivity increases during waking and decreases during sleep. This occurs in thalamocortical areas, i.e., in those areas of the brain in which slow-wave activity is generated. Functionally, the theory explains why sleep has a restorative function while also enforcing learning and memory: sleep would not only down-regulate synaptic strength in general, thereby saving the energy needed to maintain the functional connections, but it would also enforce the differences in synaptic connectivity that originated from learning processes and activity during wakefulness. The theory also explains why slow wave activity can be enhanced in specific areas of the brain by selective, increased use of that area [36–38]: in this way synaptic connectivity is linked to anatomy.

3.2. Process C

During the last decades, knowledge of the circadian system has increased enormously. In the 1970s and 1980s, the theory emerged that the circadian system would consist of a single pacemaker in the brain, telling time to all kinds of other non-rhythmic processes [39]. At that time it was already clear that this theory did not hold for every process (with the so-called food-entrainable oscillator [40] as a clear example of an exception). By now it is known that many of those other processes are autonomously rhythmic [41]. These so-called peripheral oscillators are not merely responding to the master pacemaker in the SCN, but also responding to other zeitgebers, specific to the function of the organ under consideration [42].

Also at the level of the SCN itself, views have changed. The SCN used to be considered a single circadian oscillator, operating as a single entity. Today we know that the circadian...
pacemaker consists of many pacemaker cells with their own intrinsic periods, the interaction of which determines aspects of the ultimate circadian signal [43–45]. In fruit flies it has been demonstrated that specific subgroups of pacemaker cells exist, one subgroup involved in the regulation of morning activity and one in the regulation of the fly’s evening activity [46–48]. The “morning cells” seem to send a daily signal to reset the “evening cells”, with no apparent interaction at intermediate times [46]. Anatomical [44], physiological [49,50], and behavioral data [51] strongly suggest that similar structures exist in mammals. Functionally, a clock system composed of a morning and an evening oscillator would be well suited to perform tasks such as tracking dawn and dusk. Anatomically, much of the fine structure of morning and evening oscillators remains to be discovered, however. If the morning and evening oscillators do have an anatomical basis, they will each have their own characteristics. These characteristics and the mutual relationships between the oscillators may explain individual differences in sleep duration and also in sleep timing, by differentially affecting sleep onset and sleep offset.

Light is the main zeitgeber to synchronize the master circadian oscillator and it was for the timing of sleep that the first phase response curve in humans to single bright light pulses was made [52]. Apart from the appropriate timing of light exposure for inducing shifts in the sleep–wake cycle, the intensity of the light also appears to be important for sleep regulation. This is particularly clear under circumstances in which subjects are generally exposed to low intensities of light; Alzheimer patients in nursing homes are among such subjects. They frequently develop sleep disturbances to the extent that they wander about at night and nap during the day. Increasing the amplitude of the light–dark cycle by increasing daytime light intensity in the ward has a positive effect on consolidating sleep during the night and wakefulness during the day [53]. Apart from direct effects of the accentuated light–dark cycle on aspects of behavior, it is possible that this is due to a regained stronger output signal of the circadian pacemaker.

In the last few years, knowledge of the pathways along which light influences the circadian pacemaker has increased remarkably. It has been demonstrated that those pathways do not rely exclusively on rods and cones, but that a third type of photoreceptor cell is involved, which is localised in the ganglion cell layer of the retina [54,55]. Those photoreceptor ganglion cells contain melanopsin as the relevant photopigment [55,56]. The cells have extensive dendritic fields, covering large parts of the retina. Thereby they are no longer suitable for image formation, but instead integrate light exposure over large parts of the surroundings: they measure overall light intensity. The cells are involved in several processes in which overall light intensity is important, like the regulation of pupil size and of the phase position of the circadian pacemaker [57]. Apart from melanopsin, it has also been suggested that cryptochrome serves as a photopigment in the non-image forming visual pathway [58]. While a critical role for cryptochromes in the generation of circadian rhythmicity has convincingly been demonstrated [59], the question whether cryptochromes serve as photopigments in the retina still needs further study [60].

4. Interaction between process S and process C

According to the two-process model, process S is the regulated variable. It is controlled by process C and by other influences such as conscious decisions to wake up (or stay awake), or by such influences as pain. Since the course of process S is strictly linked to sleep–wake behavior, and since the phase angle of process C is controlled by light exposure of the retina, process S indirectly influences process C. Closed eyelids reduce light intensities falling on the retina by a factor of about 30, depending on wavelength [61], so sleep modifies retinal light input and, thereby, the phase of the circadian pacemaker. The use of curtains and electric lighting also contribute to the behavioral control over the circadian pacemaker. Certain behavior, e.g. going to bed late, leads in this way to delays of the circadian pacemaker which reinforces itself and eventually may lead to large differences in the phase angle between sleep and the light–dark cycle. In humans, groups of extreme chronotype exist [62,63] with differences in sleep phase of more than 4 h under entrained conditions. For these extreme chronotypes it might well be that they are extreme chronotypes not only due to abnormalities of their biological clock [64,65], but also because the behavioral patterns resulting from the clock abnormality in turn increase the abnormality of phasing of the clock—it becomes a vicious circle.

Although controversies between the various models concerning the precise way of interaction between process S and process C remain, there are clear ideas about the location in the brain where these interactions occur. In a review in 2005, Saper et al. [66] argued that the dorsomedial nucleus of the hypothalamus is the site at which the circadian rhythms in feeding, locomotion, sleep–wake alternation, and corticosterone secretion are regulated. At this site, circadian signals coming from the SCN through the subparaventricular zone combine with inputs from other areas to allow for flexible control over sleep timing [67]. It is likely that the need for sleep—due to the prior history of sleep and wakefulness—is one of these other influences.

There is evidence that the interaction between process S and process C is much more intimate and basic than just at the level of interactions between cells of various brain areas. Such evidence arises from a completely different field of knowledge: molecular and genetic mechanisms for the generation of circadian rhythms. A series of mammalian circadian clock genes has been discovered. Together with their protein products, those genes are part of transcriptional/translational molecular feedback loops, and these loops underlie cellular circadian rhythmicity [68]. Interestingly, some of these clock genes also seem to play a role in sleep-regulatory processes; in both Cry1Cry2 double knockout mice and BMAL1/Mop3 knockout mice, not only are the circadian rhythms of locomotor activity disturbed but also the amount of sleep (and of SWA) is increased, sleep consolidation is disrupted, and the response to acute sleep deprivation is attenuated [69,70]. It is hard to understand these effects on the basis of changes of the circadian pacemaker alone, suggesting the presence of a molecular basis for the interaction of homeostatic and circadian processes in sleep regulation.
5. Concluding remarks

In virtually all organisms, the predictability of the temporal changes of the environment has stimulated the development of a circadian pacemaker, which is used to anticipate those environmental changes. This generalisation also applies to humans. The circadian pacemaker controls all sorts of behavior; in many species this includes the timing of sleep. As long as the predictability of the environment continues, such a system is very advantageous because it increases the efficiency of all processes that need to be activated in rhythmically changing environmental situations. However, the availability of electricity has reduced enormously the predictability of our environment. We can be active at will at any moment of the day. The contemporary 24-h society that operates 7 days each week almost forces us to do so. Since the biological system has not evolved under industrial conditions requiring human effort around the clock, it is not optimized to such conditions. Evolutionary processes develop slowly, and it may take many generations before our circadian system can cope with its new and unpredictable environment, if this is possible at all. Until such a time, we can only help our own performance and well-being by investigating the detailed mechanisms that underlie the circadian control of the timing of sleep and by using that knowledge to trick the system towards optimal performance.

Acknowledgment

Our work is supported by the EU 6th framework project “EUCLOCK” (FP6 Grant 018741).

References


